Appendix D:
Emissions Data Compilation, Editing, and Reduction
and
the Analysis of Variance Approach to Statistical Treatment of Emissions Data

Raw data files of the emissions tests from each laboratory were submitted electronically and then loaded into the Alternative Fuels Data Center at NREL. Before any data analysis was conducted, checks and edits were undertaken to ensure data quality. In particular, the data were reviewed for the presence of outliers. To begin this review process, the data sets were sorted by vehicle type, test fuel, and test round. At the first level of data quality checks, the replicate test results were evaluated. An initial set of replicate tests was conducted on some vehicles to provide information about test repeatability. Additional replicated tests were performed on vehicles that exceeded the EPA emissions certification standards. A comparison of the replicate results helped to identify some individual test results as outliers. These results were then eliminated from further consideration (although, as described below, the established outlier detection procedure involved more than these replicate test results).

The four-stage procedure outlined below was used to identify and eliminate outliers in the exhaust emissions test results, and to compile the final data sets for statistical analysis. No evaporative emissions results were removed from the data sets because of the high level of variability in typical evaporative emissions.

1. Stage One (Replicate Analysis)—For each emissions constituent (e.g., NO_x), all pairs of replicated test results were first considered. The absolute value of the difference between each pair was computed, and the mean and standard deviation of all such differences were also computed. Individual differences outside a bound equal to the mean plus three standard deviations were flagged as excessive. The two test results from each of the flagged pairs were then reviewed, and the one result in each pair furthest from the overall mean was designated as an outlier and eliminated. For all other pairs (those not flagged as excessive), the two test results were simply averaged to produce a single result. In this manner, the overall data set was reduced to a single value per vehicle type/fuel/test round for each emissions constituent.

2. Stage Two (Among-Vehicle Data Quality Checks)—Having a single set of values for each vehicle type/fuel/test round, it was then necessary to compare the results for each combination of the three (e.g., Dodge Spirit, M85, round 1). Consequently, for every vehicle type/fuel/test round combination, the mean and the standard deviation of each emissions constituent were computed. Individual vehicle values outside a bound of the mean plus or minus three standard deviations were designated as outliers and removed from further consideration.

3. Stage Three (Checks Among Emissions Constituents, or Total Vehicle Viability)—Depending on the emissions constituent in question, the application of the edits performed in Stage Two left a number of “holes” in the data. In some cases, the process resulted in multiple holes (more than one emissions constituent missing) for a given test. Because each hole is the result of an emissions test value being designated as an outlier, tests (for a given fuel/test round combination) having two or more holes on major emissions constituents (HC, NO_x, and CO) were deemed to be “not viable” and were completely eliminated from further consideration.

4. Stage Four (Data Reduction for Multiple Rounds)—Finally, for purposes of this particular report, only the results on vehicles tested in all rounds (for a particular model/fuel combination) were retained for data analysis purposes (Note: some vehicles were not tested in all rounds for a number of reasons. For example, some failed the pre-test maintenance checks and were returned to the agencies, and some were retired from service by GSA before all rounds of testing could be completed).

Analysis of variance (ANOVA) was the principal statistical technique used to analyze the emissions data presented in this report. Whereas the t-test—one of the most frequently applied statistical procedures—is used to assess the significance of differences in pairs of mean values, ANOVA facilitates simultaneous assessment of multiple differences among a collection of two or more means (see, for example, Table D-1).

Table D-1. Example Table of Mean Values

	Round 1	Round 2	Round 3
Fuel 1	$\_{11}$	$\_{12}$	$\_{13}$
Fuel 2	$\_{21}$	$\_{22}$	$\_{23}$

Note: See below for explanation of “fuel” and “round.” _ stands for the mean value of some emissions constituent of interest (e.g., CO). $\_{11}$ - $\_{23}$ is an example of one possible difference in mean values.

ANOVA is even more useful in that it allows the total variation in a set of data (as measured by the sum of squared deviations from a mean value) to be subdivided into the portions that are attributable to various experimental or observational factors. In this manner, the contributions of various factors to the observed variability in some test result, laboratory response, or property of interest, can be identified and quantified, along with the effects of such factors interacting among themselves.

In the context of the emissions testing program discussed in this report, the experimental factors assumed to generate differences in test results are: (1) fuel (alternative fuel versus gasoline); (2) round (a proxy for mileage); (3) laboratory (three different laboratories chosen through competitive bidding and employing the same test procedures; one of the three at high altitude); and (4) vehicle model (Dodge Caravan, Chevy Lumina, etc.). In addition, differences among individual vehicles of the same model contribute to the total variation in emissions test results, with random sampling resulting in such differences. Although other factors may affect variability in emissions, these are not explicitly controlled in the test program. Contributions to the total variation from these factors cannot be determined.

The arithmetic computations of analysis of variance, which are explained in textbooks on statistical methods, are usually summarized in a tabular form like the one shown in Table D-2. The first column in the table identifies the experimental factors, or sources of variation, while the second lists the corresponding numbers associated with a quantity called the “degrees of freedom.” Typically, the degrees of freedom associated with a particular factor consist of the number of “levels” of that factor minus one (or in the case of the category labeled “Total,” the overall number of observations or test results minus one).

Table D-2. General Form of an ANOVA Table

Source	Degrees of Freedom	Sums of Squares	Mean Squares	F-Value	Significance Level
Total	n-1	*			
Factor A	a - 1	*	*	*	*
Factor B	b-1	*	*	*	*
. . .	. . .	*	*	*	*
Factor Z	z-1	*	*	*	*
Remainder[1]	(n-1)-(a-1)-(b-1)-…-(z-1)	*	*		

*Values to be computed.

[1]In many cases, "Remainder" is denoted as "Error," which, depending on the context of the analysis, can be either experimental error or sampling error.

Note: n is the total number of observations; a is the number of levels of Factor A, b is the number of levels of Factor B, etc.

The third column lists a series of intermediate calculations, referred to as "sums of squares," which are associated with the respective factors or sources of variation. "Sums of squares" is abbreviated wording for "sum of squared deviations from the mean," which is the basic calculation needed for computing a statistical variance. The sums of squares associated with the different factors in Table D-2 below the "Total" line must, of necessity, add up to the sum of squares shown on the "Total" line (this is the additive property of ANOVA).

The fourth column in Table D-2 lists a series of numbers referred to as the "mean squares." The mean squares associated with the respective factors or sources of variation are computed by dividing the corresponding sum of squares by the corresponding degrees of freedom. It is these mean squares that are actual variances.

The fifth column in the table contains a series of numbers under the heading of "F-Value." These numbers are determined by taking ratios of the mean squares associated with various factors. The numbers in this column are referred to as F-values because they adhere to a special probability distribution called the F-distribution.

The sixth and final column in the table lists probability values that can be used to assess the size of the corresponding F-values (or ratios of "mean squares"). These are often referred to as "Significance Levels."

Typical ANOVA tables based on some of the data presented in this report is shown in Tables D-3 and D-4.

Once the experimental factors, or sources of variation, have been accurately identified, the calculations necessary to complete an ANOVA table are relatively straightforward. Software products such as JMP, available from SAS Institute, make it possible to avoid the algebraic tedium that would otherwise be required to compute all the numbers. Interpreting the results is quite a different matter. To make an appropriate interpretation, we must consider the population of units to which statistical inferences are to be drawn. In addition, we must determine which factors are to be regarded as "fixed" and which are "random."

Table D-3. ANOVA in CO Measurements Obtained in Emissions Tests on Flexible-Fuel Dodge Intrepids

Source	Degrees of Freedom	Sums of Squares	Mean Squares	F-Value	Significance Level[4]
Total	59	1.5431			
Rounds	1	0.2926	0.2926	10.1715	0.0066
Fuels	1	0.0150	0.0150	1.1894	0.2939

Round x Fuel[1]	1	0.0023	0.0023	0.2306	0.6385
Vehicles	14	0.5149	0.0368	1.1669	0.3941
Vehicle x Round[2]	14	0.4027	0.0288	2.9077	0.0275
Vehicle x Fuel[3]	14	0.1771	0.0127	1.2783	0.3261
Error	14	0.1385	0.0099		

[1,2,3]Factor interaction terms
[4]Values of .05 or less would ordinarily indicate significant differences. For example, the significance level of 0.0066 associated with the F-value for "Rounds" indicates that the same average value of CO was not obtained in both test rounds.

Table D-4. ANOVA in NO_x Measurements Obtained in Emissions Tests on Flexible-Fuel Dodge Spirits

Source	Degrees of Freedom	Sums of Squares	Mean Squares	F-Value	Significance Level[4]
Total	83	1.2592			
Rounds	1	0.0074	0.0074	0.4901	0.4920
Fuels	1	0.0031	0.0031	0.2444	0.6264
Round x Fuel[1]	1	0.0500	0.0500	17.5905	0.0004
Vehicles	20	0.5851	0.0293	1.1708	0.3381
Vehicle x Round[2]	20	0.3032	0.0152	5.3304	0.0002
Vehicle x Fuel[3]	20	0.2534	0.0127	4.4551	0.0008
Error	20	0.0569	0.0028		

[1,2,3]Factor interaction terms
[4]Values of .05 or less would ordinarily indicate significant differences. For example, the significance level of 0.0004 associated with the F-value for the "Round x Fuel" interaction indicates that the difference in the average values of NO_x for the two fuels was not the same from one test round to the next.

Fixed factors are those whose range of values, or levels, are completely encompassed by the specific population units included in the investigation. In the context of the this emissions testing study, "fuel" is a fixed experimental factor because there is not interest in, nor rationale for, drawing conclusions about fuels other than those being specifically studied. A random factor, on the other hand, is one about which conclusions can be extended to a larger collection of units than the ones specifically included in the investigation. In this context, "vehicle" is a random factor because individual vehicles were randomly selected from a larger collection, or population, and projecting the results of the testing program to that larger population is desirable. The determination of fixed and random factors governs the way the F-values are computed (that is, the choice of numerator and denominator in the ratio of mean squares; the denominator always represents an "error" term against which the numerator is compared) and directly affects interpretation of the results. The bigger the F-value, the more likely at least one difference among the means being compared is statistically significant.

ANOVA's statistical procedure is constructed on certain mathematical assumptions. The first assumption—that effects of the various experimental factors are additive—has already been mentioned (in the sense that the individual sums of squares add up to the total). The second assumption is that all experimental errors are random, independent, and follow a normal (Gaussian, or bell-shaped) distribution. Violating either of these assumptions will negate the interpretability of the results.

Statistical software packages such as JMP provide many other capabilities that extend and build on the information derived from the basic ANOVA. In particular, it is possible to estimate the actual components of variance attributable to each experimental factor, and to adjust mean values for unequal numbers of observations using a least squares approach. The details of these techniques are beyond the scope of this discussion.